01

II

Part no. 2, 3rd QTR

7.6 - 10.5 00

CR-148820

AGRESTE

(E76-10500) AGRESTE PROGRAM. PART 2: FRENCH TEST-SITES Progress Report, Mar. - Jun. 1976 (Commission of the European Communities) 29 p HC $4.00 CSCL 02C

N76-32613

Unclas
G3/43 00500

PROGRAMME

PROGRESS REPORT

MARCH - JUNE 1976

28790

PART 2

FRENCH TEST-SITES

Original photography may be purchased from:
EROS Data Center
10th and Dakota Avenue
Sioux Falls, SD 57198

C O N T E N T S

CHAPTER I

A FAST CLASSIFICATION METHOD-APPLIED TO RICE-FIELDS IN THE CAMARGUE TEST-SITE (SOUTHERN FRANCE)

T. LE TOAN

Centre d'Etude Spatiale des Rayonnements
TOULOUSE

Abstract

A study of the Camargue rice field responses from Landsat imagery is made, taking into account the ecological and agricultural conditions of the zone.

The large dispersion of the rice-fields responses in comparison with other types of ground objects, permits to develop a fast method of classification using the linear classifier.

The results are compared with that obtained by a conventional method.

I - Description of rice parcels in the Camargue

The arrangement of rice fields in the Camarque (Southern France) depends on the ecological and agricultural conditions of the zone.

Areas of rice cultivation consist of adjacent parcels whose dimensions depend mainly on wind conditions and topography.

The parcels are leveled so that the depth of the water can be controlled to within one centimenter. Prior to levelling, a topographical study of the terrain is made. The parcels are delineated in such a way to minimize the volume of earth to be removed and to reduce the harmful effect of soil rasping. This implies that the parcels be narrow and that their length follow contour lines.

Furthermore, since the prevailing wind, the Mistral, which causes damage particularly during the first months of vegetation, blows North-South, the fields are narrow and orientated East-West in order to reduce the action of the wind.

Given these restrictions, most of the fields have a small area of from 1 to 3 hectares and are rectangular.

Figure 1 shows a view of the Camargue rice fields taken on June 20 1975 by Daedalus multispectral Scanner at 7000 m (near Infrared Channel : 0,900 - 1,100 μm)

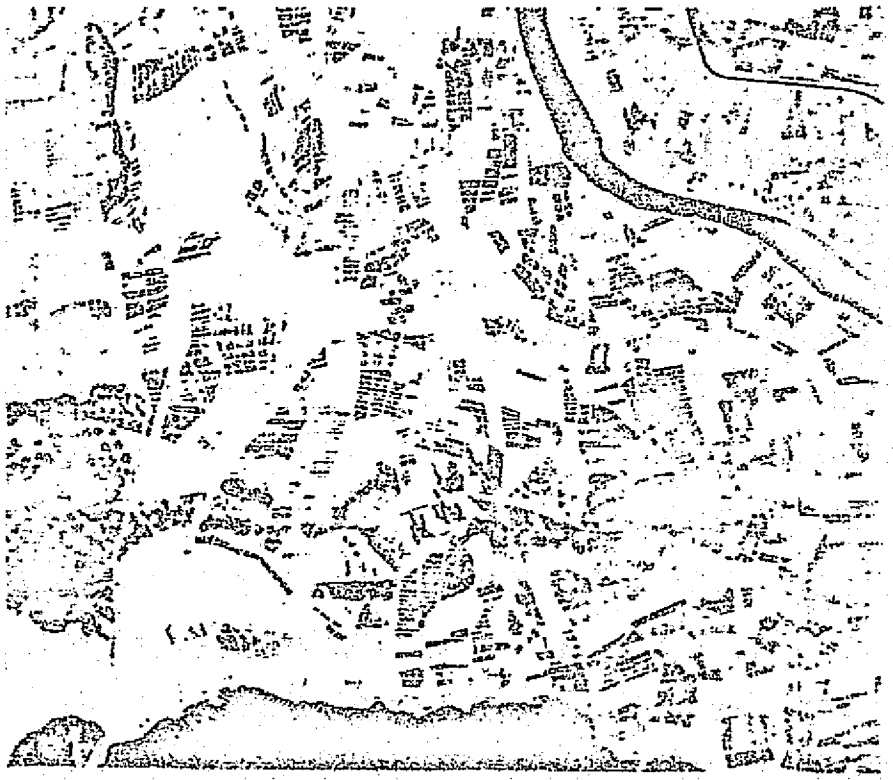


Figure 1

View of the Camargue rice fields
Scale : 1/100.000

The fields are separated by dirt roads which must permit transportation operations (sowing, chemical treatments, harvest...) which often take place when the fields have been flooded. These roads are from 4 to 10 meters wide and take up from 15 to 25 % of the total cultivated area.

.../...

After flooding, the fields look line a mosaic of rectangular surfaces of about 1 to 3 hectares, with the water level at about 10 cm.

During sowing, the layer of water serves as a temperature control for the shoots. Thus the incessant motion of the water helps to compensate for variations in air temperature.

From August to September, the movement of the water (emptying and refilling) depends on evaporation, losses through drainage and the water need of the plants.

Therefore, irrigation condition differs from a parcel to another.

Furthermore, during the vegetative stage of rice, the coverage of the vegetation on parcels at a given time varies from parcel to parcel, depending on the variety planted.

II - Study of the response of reference rice-fields

The reference rice fields located within a block of parcels by means of ground observations and aerial photography are selected from Landsat 2 recordings.

Each pixel on Landsat image either coincides with one parcel or straddles two.

The homogeneity of the rice class is studied by means of the histograms of the responses of the selected samples.

A zone is defined as homogenous when the gradient at all points in the zone is zero or below a threshold determined by the measurement noise.

The gradient can be calculated following 8 directions. This leads to consider zones of 3 x 3 points (or pixels).

.../...

$$\begin{array}{ccc} \underset{\bullet}{U_{11}} & \underset{\bullet}{U_{12}} & \underset{\bullet}{U_{13}} \\ \underset{\bullet}{U_{21}} & \underset{\bullet}{U_{22}} & \underset{\bullet}{U_{23}} \\ \underset{\bullet}{U_{31}} & \underset{\bullet}{U_{32}} & \underset{\bullet}{U_{33}} \end{array}$$

If $|U_{ij} - U_{22}| \leqslant \alpha$
$i, j = 1$ to 3
with α = noise threshold

the zone of 3 x 3 points is considered to be homogenous.

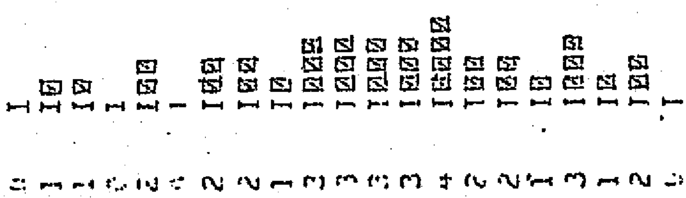
In addition to this concept of instrumental homogeneity there is also the concept of the homogeneity of the ground object class. The latter concept is related to problems of resolution.

With LANDSAT resolution (80 m x 80 m) such agricultural terrain as vineyards or orchards (with a structure of alternating vines and soil or trees and soil) appears as stable composite and the samples taken within this type of terrain have homogenous responses.

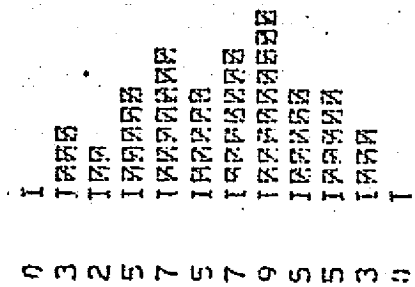
However, for rice-fields in the Camargue, which consist of small parcels with differing states, differences in irrigation conditions and in rice varieties, the responses of the pixels taken within a sample area give a greater-dispersion, especially in the MSS7 channel.

Figure 2 gives 2 examples of histograms for 2 of the most important types of agricultural terrain in the Camargue, rice fields and vineyards

...



rice fields



vineyards

Figure 2

Histograms of rice-field and vineyard samples.
(MSS7 responses - July 6 1975)

6 - Method of classification

The fastest method for mapping of rice fields from data obtained from May to July, when ricefields are flooded or in early vegetative stage consists of density slicing of the MSS7 image.

When data are acquired during the vegetative stage of rice, i.e. from July to September, responses in MSS7 channel of rice-fields overlap that of other crops (corn, cereals), as it is shown in the figure 3.

In the figure 3, three classes of objects are studied by a bispectral diagram of the responses of samples in channels MSS5 and MSS7 :

rice fields, vineyards and cereals (corn, wheat) from the July 6 1975 scene (2165-09511).

The average variances in responses and the correlation coefficients for the 3 classes are the following :

	σ_5	σ_7	r
Rice field	4.37	6.44	-0.117
Vineyard	4.48	3.57	-0.012
Cereals	7.57	5.15	-0.005

For the calculations, radiometric value is coded with 8 bits (from oto 255) When studying the diagram, we can make following remarks :

- The projection of the rice-field class (1) on MSS5 axis will be merged with that of the vineyard class (2) : same density range in the chlorophyll band.

- The projection on MSS7 leads to an ambiguity between class 1 and class 3 (cereal) : existence of several irrigated cereal fields.

This density slicing applied on MSS7 image will give rice field class mixed with cereal class.

On the other hand, we can see that the plotted points which correspond to pairs of measurements x_i, y_i in MSS5, MSS7 bands lie along an axis. For each class, the slope of the axis depends on the correlation between x_i, y_i. If the characteristics were completely independent of one another, the mean value of their product would be 0, and the axes would be axes of x and y.

In this example for class "2" and class "3", the MSS7 responses are rather independent of MSS5 responses. This means that the correlation coefficient, which is :

$$r = \frac{\Sigma\, x_i\, y_i}{\sigma x . \sigma y} \quad \text{is close to 0}$$

.../...

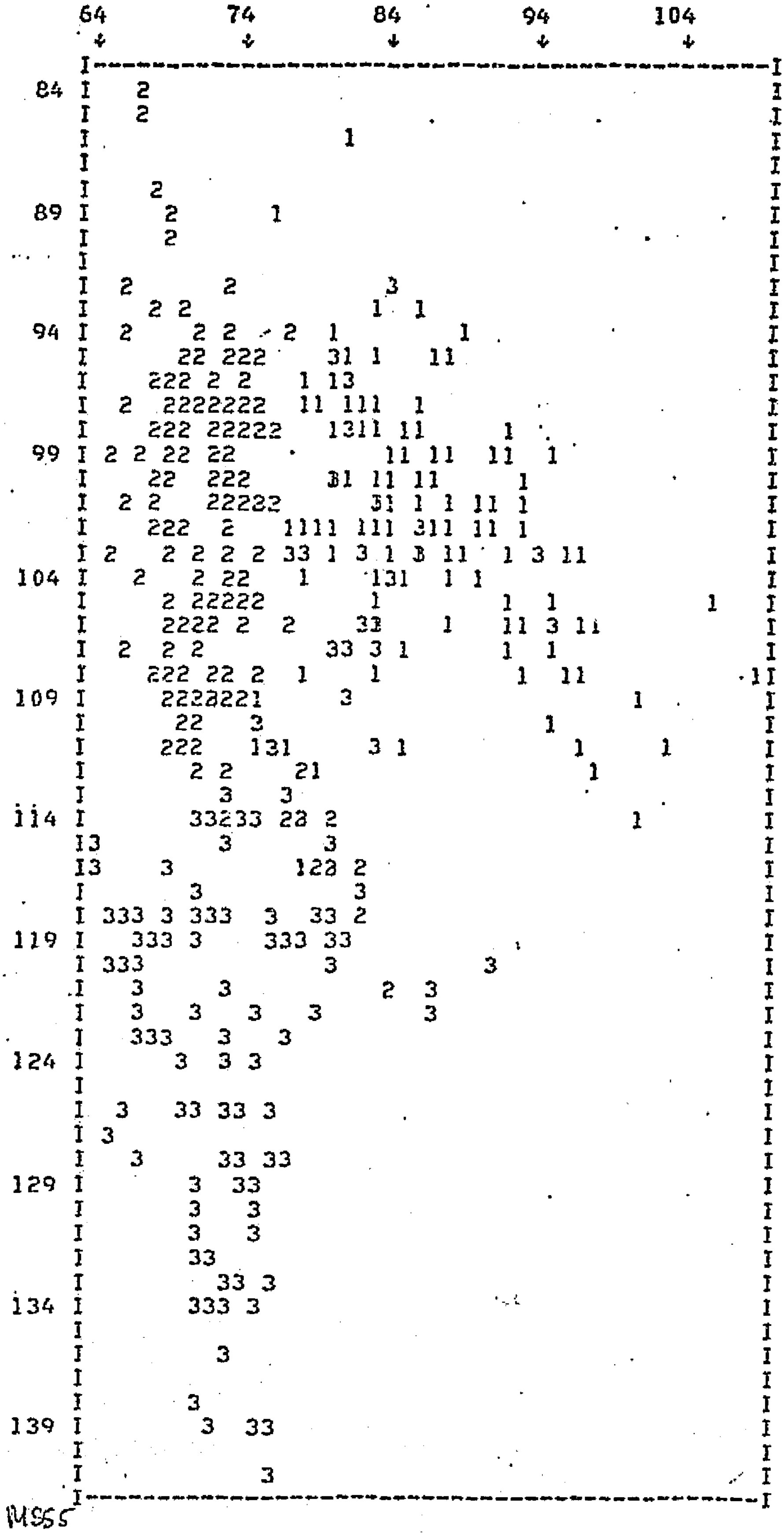


Figure 3

Bispectral diagram of the responses of samples(MSS5 and MSS7 bands)
1 : ricefield, 2 : vineyard, 3 : cereals.

The slope of the axis can be calculated, if we assume that the cloud-formed by plotted sample points of every class is an ellipse, by : *

$$tg\ 2\theta = \frac{-2r.\sigma x.\sigma y}{\sigma x^2 - \sigma y^2}$$

with θ : angle between x axis and the ellipse great axis.

Since r is close to 0, either θ will be very small, or else σx is nearly equal to σy.

In this case, where $\sigma x \neq \sigma y$ for the 2 classes, the possible conclusion is that θ is very small, and the axes of class "2" and class "3" clouds nearly coincide with the x axis.

On the other hand, when we compare the slopes of the axes of class "1" and class "2", as

$\sigma x_1 \simeq \sigma x_2$ and

$\sigma y_1 > \sigma y_2$:(dispersion of rice-field responses in MSS7 band)

since $r_1 \gtrsim r_2$)

$tg2\theta_1 > tg2\theta_2$

The slope of the rice field class axis is greater than that of the vineyard class.

Then, the fastest method of classification consist to choose a linear classifier in MSS5 _ MSS7 plane. The equation of the line can be easily determined on the diagram.

The plane is then separated into 2 regions : region of the ricefield class and the region of other classes. We must however to ascertain that the region 1 does not include other classes (water bodies...).

* The calculations and notations are from = An Introduction to Mathematical Probability. Julian Lowell Coolidge - Dover publications, INC. NEW YORK.

.On other example of the application of the method is shown in the figure 4, where the plot of the bitemporal data permit to separate ricefield class from the others by a linear classification.

IV - Result

Figure 5 shows the result of classification using linear classifier from the scene of July 6 1975. For 180 x 180 pixels (15 km x 15 km) this method gives 45 % area occupied by the ricefield.

The calculation time is 0.07 sec CDC 7600.

Figure 6 shows the results of classification method using Euclidian distance. On the same surface, the results give 41,3 % ricefields, 20 % vineyards, 29 % areal and corn. The cal ulation time is 1,5 sec CDC 7600.

For both results, the reference ricefields which are : Clos des Saules, Mas des Aubépines, Mas d'Adrien, Mas de Vert are well defined.

In figure 7 the outline of a ricefield, the Clos des Saules, obtained by processing of LANDSAT imagery is compared with a ground survey map. The overall area of the field is found on the LANDSAT results, but in fact the roads and dikes within a block of parcels occupy a non productive area of from 15 to 20 %, according to investigations made on the 3 parcels : les Pébrières, le Clos des Sàules and le Mas de Vert.

On the result of classification in figure 6, a smoothing technique is applied. For every pixel to be classified, the 8 surrounding pixels are considered. The smoothing technique takes into account the distance between pixels, the class of the central pixel and that of the surrounding pixels.

This smoothing technique which gives good results in large homogenous zones is considered unsuitable for rice fields in the Camargue for identification of blocks of parcels. The figure 8 shows a degradation in the contour of the small area rice fields : this is indeed the case of the Mas des Saules, Mas des Aubepines, and the Mas d'Adrien.

9a

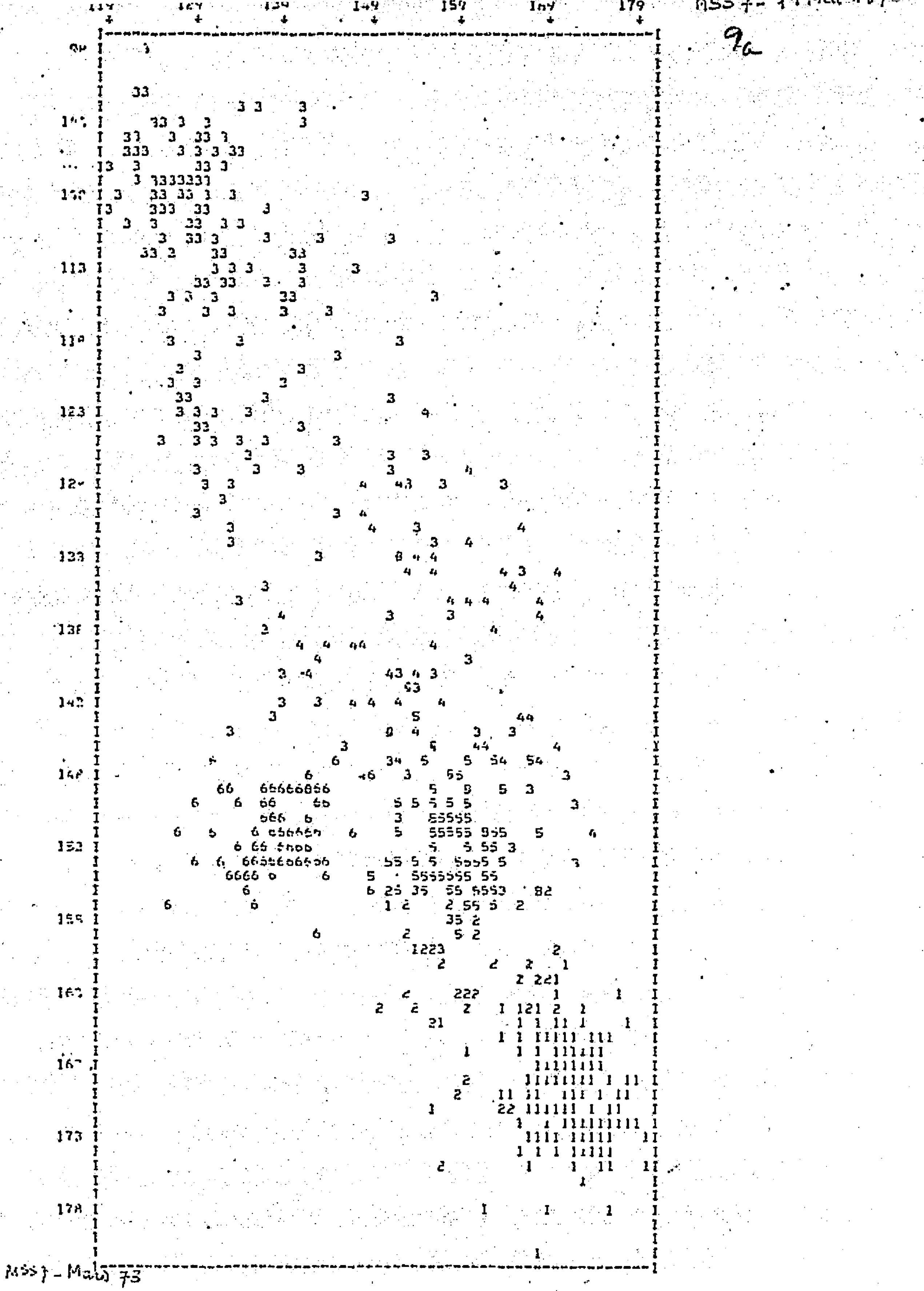


Figure 4

Diagram of data from MSS7 band in 2 dates : March 1973 and May 1973

Class 1 : meadow and cereal, 2 : vineyard, 3 : Rhône, 4 : petit Rhône, 5 : urban, 6 : ricefields

Figure 5

Classification of July 1975 scene using linear classifier

9c

Figure 6

Classification of the July 6 1975 scene using Euclidean distance

dark : rice

medium : vineyard

light : cereal

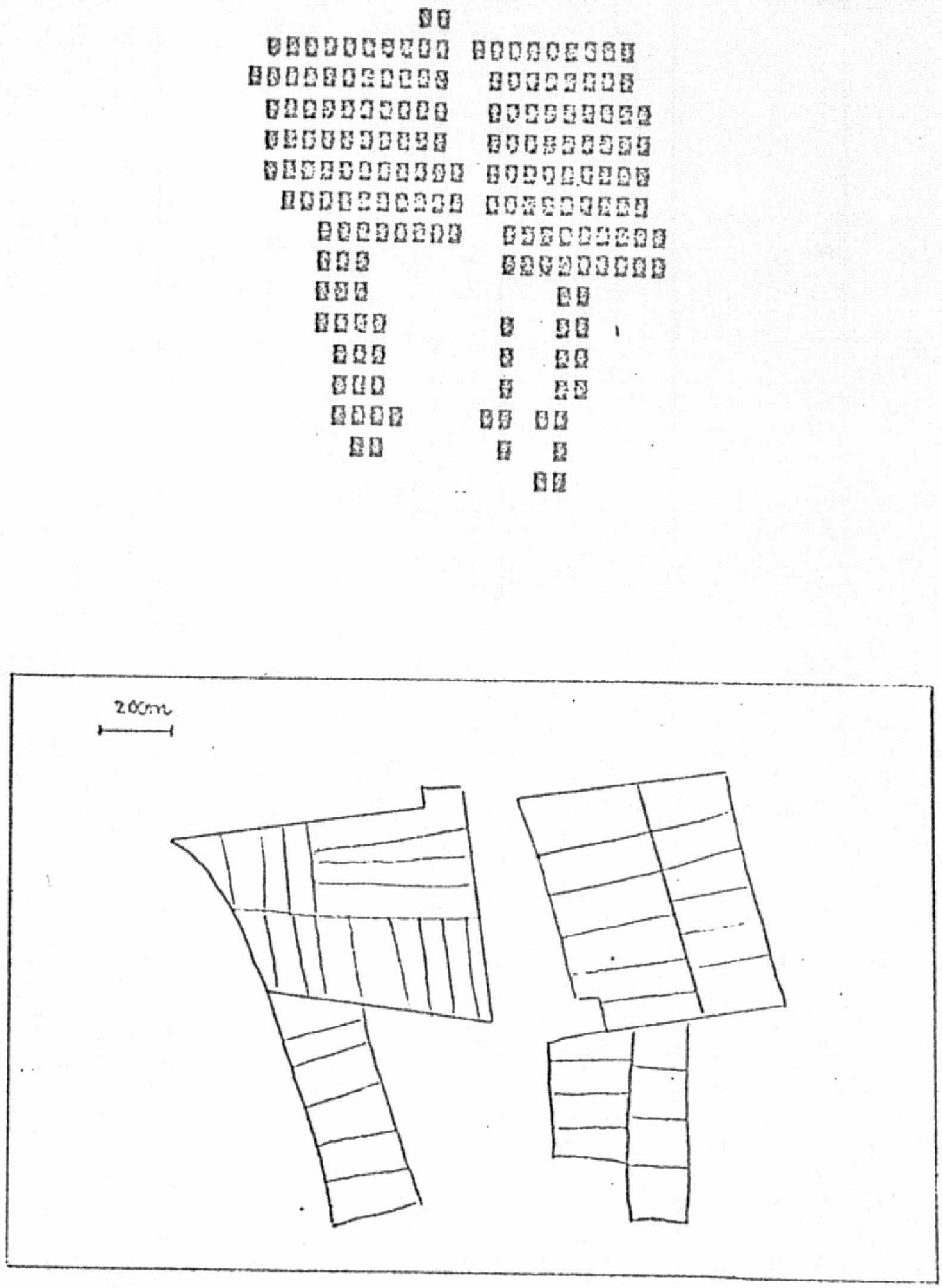


Figure 7

Comparison of Landsat result and ground survey mas of 2 reference ricefields.

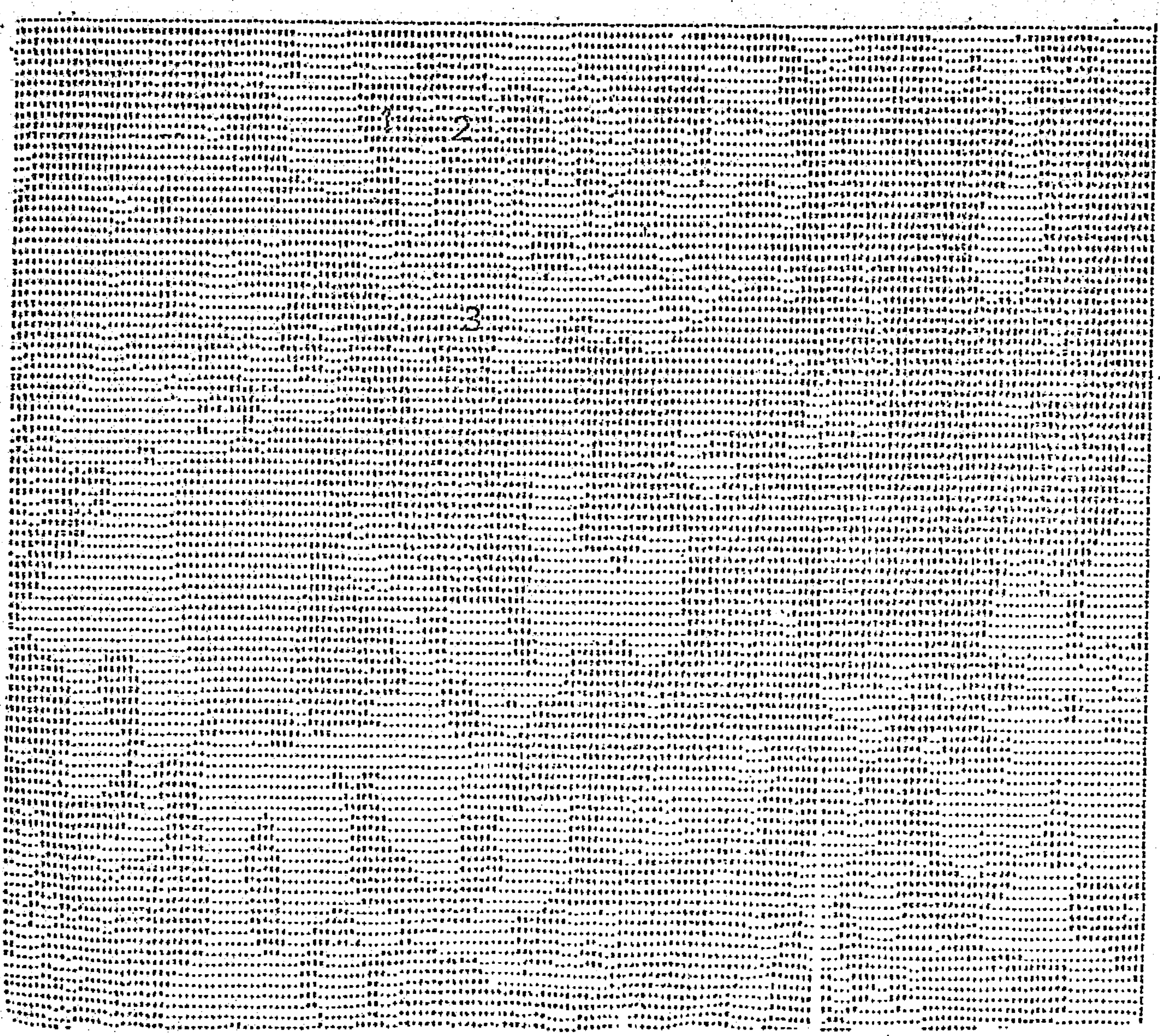


Figure 8

Result from classification method on July 6 1976 imagery followed by smoothing technique

1 - Clos des Saules 2 - Mas des Aubepines 3 - Mas d'Adrien

CHAPTER 2

2-1

Unsupervised classification method applied to rice-field characterisation from LANDSAT 2 imagery

P. CASSIRAME, T. LE TOAN
Centre d'Etude Spatiale des Rayonnements

Abstract

The mobile centers clustering technique used by the C.E.S.R. as unsupervised classification method is described and especially the initialisation of the process which is carried out in such away to reduce the processing time and the memory needed, is explained.

The application of the method for ricefield mapping from LANDSAT data is shown, and the results are compared with that obtained by supervised techniques.

The evolving of supervised methods is based on the existence and the properties of prototype elements which correspond to ground objects, determined by ground investigations.

Whenever prototype samples are not acquired, unsupervised methods must be developed.

The technique evolved by the C.E.S.R. is a "mobile centers" clustering technique called "Nuées Dynamiques" (Dynamic Clusters) by this author, E. DIDAY (1).

I - DESCRIPTION OF THE METHOD

I.1. Principle

Given a metric space (E,d) and a set $k \in \mathcal{P}(E)$ called the set of centres.

To k is associated a partition of E into polyedral, convexe subsets Ec :

$$E_c = \left\{ x \in E \;/\; \forall c' \neq c \in k \;:\; d(x,c) < d(x,c') \right\}$$

Each set Ec has a center of gravity Gc and for each Ec, a cluster made up by points Nc called nuclei is built. Nc set is composed by points condensed aroung Ge and is determined by its cardinal nc :

$$N_c = \left\{ x \in E_c \;/\; \mathrm{Card}\ N_c = n_c ,\ \forall x' \notin N_c : d(x,G_c) < d(x',G_c) \right.$$

The center of gravity of Nc cluster is then calculated $\rightarrow G_c^1$ and another set of centre K_1 is defined. The elements of which are the nearest elements around the new centre of gravity G_c^1.

A new partition of E into subset E_c^1 -center G_c^1- is then constructed as above and the subset Gc is then superseded by G_c^1.

.../...

The dynamics cluster method consists then to iterate the process, i.e. to build new centers Gc, G_c^1, G_c^2..., G_c^n and to stop at the iteration i if $G_c^i = G_c^{i+1}$ $\forall$c.

The method is justifies by the fact that the passage from G_c to G_c^1, G_c^2, ... reduces the intra-classes variance of Ec.

I.2. Convergence of the process

If the cluster E is finite, there is then a finite number of partitions of E. To each partition corresponds a total internal variance that decreases strictly after every iteration.

The variance will reach a minimum after a certain number of iterations. Further, successive iterations modify only boundary points, i.e. the when the distances from these points to 2 distinct centers are equal.

Remark 1 : Instead of the center of gravity G_c of N_c set an element of N_c set can be taken. The creation of any artificial non convexe class is then avoided.

Remark 2 : E. DIDAY advises to use all the points of the N_c set to build the new partition E_c^1. That allows :

- to reduce the effect of the marginal elements.
- to apply various metrix, for example the Mahalanobis metrix which cannot be used when the N_c set is reduced to one point.
- to avoid the clustering of the 2 points far from each other but linked by a line of points.
- to use the totality of the samples if the method is used in supervised manner.

I.3. Initialisation of the process

The determination of the set of centers, k, can be made :

- either in studying the results from a correspondence analysis and that needs a supplementary treatment.
- or randomly, and that leads to a great number of iterations

.../...

To avoid these 2 inconvenients, the technique carried out by the C.E.S.R. is the following :

The area under study is subdivised into sub-areas Z_1, Z_2,..., Z_n, each of which is a square of 60 x 60 pixels.

The process is applied first to the zone Z_1. The initial set of centers is chosen randomly and the iteration number is limited to 5. If after 5 iterations, the convergence of the process is not obtained, it is decided that there was bad initialisation, then an other random choice of initial centers is made and the process starts again.

When the convergence on Zi is obtained, sone Zi+1 is then studied with the last calculated set of centers of Zi as initial centers. As there is close neighbourhood between Zi and Zi+1 with best chance of fitting and the convergence is reached practically after one iteration.

II - APPLICATION TO AGRICULTURAL MAPPING IN CAMARGUE

The test-site is an area of 180 x 180 pixels (about 15 km x 15 km) in the region surrounding Arles (figure 1), where the ground truths are well known by ground investigations. The results from unsupervised classification can be then compared with that obtained by supervised method.

The example shown in the following concerns multitemporal Landsat data. The MSS5 and MSS7 bands from the two scenes of March 21 1973 and May 14 1973 are used.

II.1. Ground truths

The main classes existing in the area are :

- The Rhône
- The Petit Rhône
- ricefields
- vineyards
- meadows and cereals
- Arles

Figure 1

Location of the test-site

At the date of March 21, the rice-fields were ploughed, wet parcels, the vineyards were at the beginning of the vegetative state and the fields of cereals as well as meadows, formed a green carpet. However, other fields could look like ricefields (ploughed corn fields).

At may 14, ricefields are flooded, except some fields where the flooding operation was done on May 15.

II.2. Application of the unsupervised method

The area under study consists of 60 x 60 pixels subareas.

Z_{11}	Z_{12}	Z_{13}
Z_{21}	Z_{22}	Z_{23}
Z_{31}	Z_{32}	Z_{33}

The process is starded on Z_{11}, where 6 initial centers are picked randomly. To the 3600 points of Z_{11} are applied the clustering algorithm with a chosen metrix.

The metrix used in this example is the euclidean distance.

When the convergence is obtained in Z_{11}, the results of the classification for 60 x 60 points of Z_{11} is preserved for display, and the last computed centers G_c^i are then conserved to be taken ad initial centers for Z_{12}.

In the case of LANDSAT imagery, each sub-area corresponds to a zone of about 4,8 km x 4,8 km. Thus, geographically, two contiguous subareas are in most case included in a same "ecologic region". This means that the same types of vegetation, of land-use exist very probably in two contiguous subareas.

Practically, one or two iteration is necessary for computing Z_{12} with the last centers of Z_{11} as initial centers. It can occur that "landscape" changes suddenly -when the sub-area is on the boundary of 2 "ecologic regions"- Then other itererations are needed, or an choice or random initial centers must be done.

.../...

The contiguity notion can be applied in both directions line and column (geographically North South or East-West). Thus, for Z_{21}, the initial centers came from Z_{11}.

Figure (2) shows the results of unsupervised classification obtained by display on Comtal device of the 9 subareas. The six classes, represented by 6 colours, are interpreted as follows :

- Class 1 : dark blue : rice field
  2 : pink : meadow and low state cereals
  3 : light blue : ?
  4 : yellow : Rhône
  5 : green : vineyards
  6 : red : petit Rhône and/or Arles ?

II.3. Comparison of supervised/unsupervised classification results :

This map is compared then with that obtained by a supervised method using the same distance (Euclidean).

On figure (3), the following classes are taken into account, after a study of reference samples :

- Class 1 : dark blue : Rhône
  2 : pink : petit Rhône
  3 : light blue : ricefields
  4 : yellow : vineyards
  5 : green : meadow and low cereals
  6 : red : Arles
  7 : black : unclassified points

The computed map obtained by supervised method is studied and compared with ground truths.

Classes 1, 3, 4, 5 are in good concordance with the ground truths, especially rice-fields which are delimited. Class 2 (petit Rhône) is partly mixed with class (Rhône) and class 6 does not represent the urban area (Arles) as expected.

The comparison of the 2 computed maps ledd then to the following remarks concerning unsupervised computed map (fig 3).

- the ricefields are well placed but their contours are not well delimited, especially in the North West side of the test-site. The edges of the Rhône are partly computed as rice-fields.

- Meadows and cereals have likeley the same delimitation in the 2 maps.

- The vineyard class does not correspond to the totality of the vineyards in the areas and class 3 is not identified, it should correspond to an intermediate radiometric responses cluster between vineyards and meadows and cereals.

- The urban area, in both maps, cannot be isolated.

II.4. Processing time problem

The processing time, on the base of CP time CDC 7600, is 0,9 sec for the 180 x 180 pixels test-site with supervised method using Euclidean distance.

For unsupervised method, the processing time, which depends on : the applied metrix, on the number of elements of every nucleus -and on the number of iteration, is evaluated in this particular case as follows :

For nucleii of 10 elements, the processing time for one iteration is 0,25 sec CDC 7600 for the first subarea -the metrix used is the Euclidean one.

The processing time for 9 x 60 x 60 pixels is approximately :

$$0{,}25 \times 5 \quad + \quad 0{,}25 \times 2 \times 8 \quad = \quad 5{,}25 \text{ sec}$$

Z_{11} $\qquad$ Z_{ij}

5 iterations $\quad$ 2 iterations (average number) :

- when Mahalanobis or X^2 metrix is used, the processing time is to multiply by 5/4.

Figure 2

Computed map obtained by "mobile centers clustering technique"

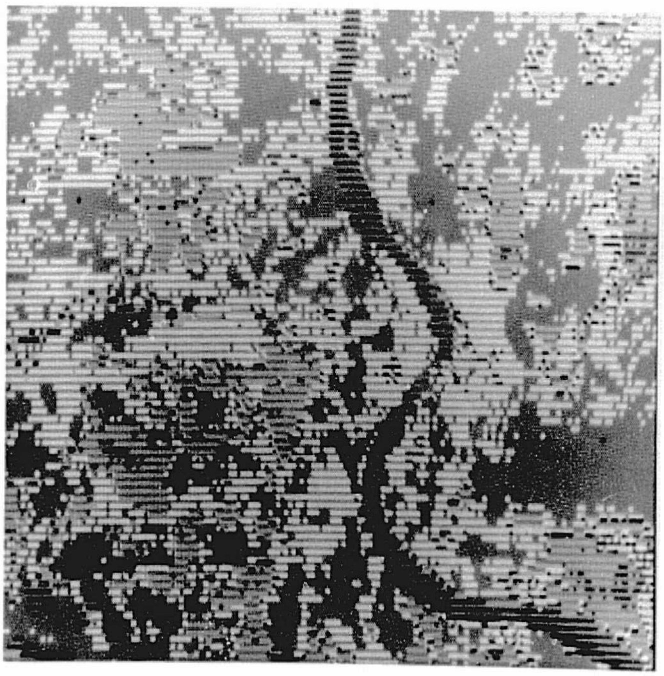


Figure 3

Computed map obtained by supervised method.